Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

STANDARD SCORES IN ARITHMETIC

S. A. COURTIS
Detroit Home and Day School

For the attainment of efficiency in any manufacturing process three things are necessary: a well-defined, a "standard," product; a knowledge of the properties and possibilities of the material to be acted upon; and a means of measuring the effects produced by the process itself. Given these three, all else is but a matter of trial and selection.

Education, from one point of view at least, is a manufacturing process. The raw material that is sent to a school is molded and shaped by the forces and experiences to which it is subjected. Marked changes are effected and the product has, in a general way, a commercial value that bears some relation to the complexity of the process and the degree of change produced. In a graded system promotion from grade to grade emphasizes the idea of "degree of change," of grade standards. Promotion from the fourth to the fifth grade, for instance, is dependent upon the acquisition of a certain amount of knowledge, the attainment of a certain degree of skill. In an efficient school one does not expect to find children of fourth-grade ability doing eighth-grade work. Each higher grade reached is supposed to mean a correspondingly higher level of ability.

The attempts of the writer to measure the efficiency of the teaching and the growth of the arithmetic classes under his control in the Liggett School have been described in some detail in previous numbers of this Journal. That work gave some knowledge of the nature of the material to be acted upon in school work, and proved that it was possible to measure accurately the changes produced by teaching effort. Clearly defined standards alone were lacking. The present article will deal with an attempt to determine such standards from the measurement of a large number of children.

Because of the complexity of the problem, the scope of the investigation was limited to the simplest work with whole numbers in the four fundamental operations. The tests used were in all essential particulars like those previously described, and will not be discussed in detail. It is sufficient to say that in their construction the utmost care was taken to eliminate the imperfections of previous tests caused by the inequalities of the units. The subject-matter of the tests, eight in number, was as follows:

Test No. 1. Addition
Test No. 2. Subtraction
Test No. 3. Multiplication
Test No. 4. Division
(Combinations, 0–9)
Test No. 5. Copying figures (rate of motor activity)
Test No. 6. Speed reasoning (simple one-step problems)
Test No. 7. Fundamentals (abstract examples in the four operations)
Test No. 8. Reasoning (two-step problems).

The nature of the investigation made it essential that the tests should be given to all grades from the lowest to the highest under identical conditions. To this end detailed instructions for every phase of the work, together with special record sheets, answer cards, etc., were printed in the form of folders to be put into the hands of each examiner or scorer. As experience had shown the most probable source of error to be faulty timing, the time allowances were printed in red. In addition a personal letter of warning was sent with each package of tests. It is realized that no degree of care can prevent either accidental errors, or deliberate fraud, but the testing of many schools tends to eliminate the effects of the first, and the character of the persons who took part in the work makes the second improbable. A few cases of accidental errors were reported, and for these tests either due allowance has been made or the results rejected entirely. The writer believes the precautions taken to have been effective; and that the conditions under which the tests were given were sufficiently uniform for the results to reflect accurately the actual conditions. Check tests of several grades under absolutely uniform conditions and with mechanical timing have yielded similar results.

The co-operation of many persons was secured by letters to personal friends, by circulars, and by editorial notice in this and two other magazines. For the kindness of these notices the writer is very grateful. It not only brought the matter to a wider circle of interested persons than could have been reached without it, but it also gave the investigation a standing it would have been hard to secure in any other way. As a result inquiries were received from many parts of the United States and from several foreign countries. The total of requests and orders for tests exceeded 30,000 copies. Ten thousand sets of tests were distributed free and eight thousand more were sold at the bare cost of printing. Teachers in England and Germany were interested in the work and the tests have been, or are to be, given in those countries also.

From the above it must be evident that the undertaking owes its success to its co-operative features. While the writer alone is responsible for whatever defects there may be in the tests themselves or in the planning of the work, credit is due the many who executed the plans and bore the burden of giving and scoring the papers. The writer is very glad of this opportunity to express his grateful appreciation of the assistance rendered by all those whose generous support and careful work has made this article possible.

Fire, accidents, sickness, and unavoidable delays of many kinds have prevented the completion of the work by many to whom the tests were sent. The tabulations of this article are based upon the returns received from some sixty or seventy schools in ten different states as shown below.

CONTRIBUTORS TO STANDARD SCORES

State	City	Name
New Hampshire	Littleton	D. F. Carpenter, superintendent
Massachusetts	Dorchester	A. R. Winter, sub-master, Mather School
	Fitchburg Leominster	E. A. Kirkpatrick, head of Department of Psychology, State Normal School
New York	New York City	H. C. Pearson, principal of Horace Mann School
		Miss Margaret Eves, acting supervisor of Elementary Mathematics, Ethical Culture School
	Brooklyn	J. C. Bell, Psychology, Training School for Teachers
	Hastings-on-Hudson	W. R. Williams, principal

CONTRIBUTORS TO STANDARD SCORES—*Continued*

State	City	Name
Delaware	Dover	A. Crawford, principal
Maryland	Salisbury	W. J. Holloway, county superintendent
Virginia	Farmville	C. W. Stone, Department of Education, State Normal School Miss Mary D. Pierce, supervisor in Practice School
	Williamsburg Toano	H. E. Bennett, professor of education, College of William and Mary
Michigan	Detroit	A. Cotter, principal, Harris School W. C. Martindale, superintendent Misses Liggett, principals, Liggett School
	Kalamazoo	S. O. Hartwell, superintendent
Indiana	Marion	J. T. Giles, superintendent
	Fort Wayne	Miss Flora Wilber, principal City Normal School
	Bloomington	H. L. Smith, superintendent
Missouri	St. Joseph	Miss Fannie Brennen, principal Garfield School
Kansas	McPherson	G. G. Pinney, superintendent

In the list that follows the number of tests sent to each of the above is given that some idea may be had of the sizes of the different groups of children measured. There is no correspondence, however, between the order of schools in the two lists. The number of actual individual scores returned for tabulation was a little over 9,000 for most of the tests.

Number of tests sent to the individuals in the list above:
2500, 1600, 1400, 1000, 750, 650, 500, 380, 350, 350, 350, 300, 200, 200, 196, 150, 140, 120
Total 12,171

A record sheet for the scores of each individual measured was sent out with each set of tests. These, when completed, were returned and will be used, later in the year, as the basis of tabulations to determine standard scores by age, and the correlation between the abilities measured. For each grade, a copy of the grade record sheet was also returned. Upon this was shown the distribution of the scores of the individuals composing the grade. Tabulations of these distributions are the bases of the tables below.

For instance, returns were received from fifty eighth-grade classes (both A and B). The distributions of the classes having the highest, median, and lowest grade averages are given in Table I; also the total of the distributions of the entire fifty classes. The same facts are shown graphically in Fig. 1.

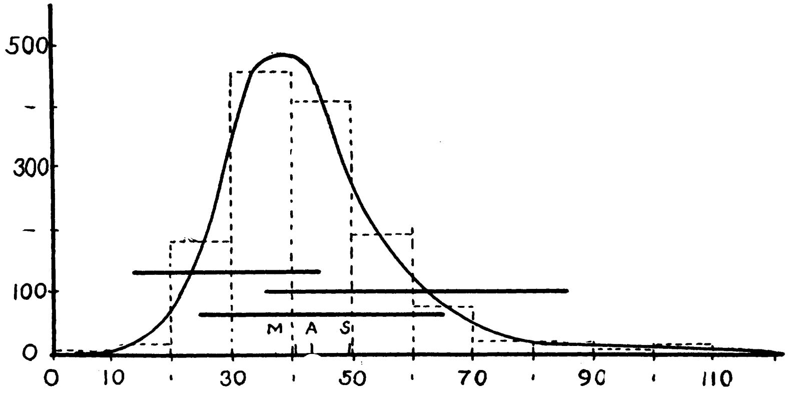


FIG. 1.—Distribution of 1,400 individual eighth-grade scores. Scale at bottom of figure gives number of answers written per minute in test of multiplication tables; scale at left gives frequencies. Solid horizontal lines show range of individual scores in the classes having the lowest, highest, and median grade averages of the fifty classes contributing the individual scores. M = median score of the entire group = 41; A = average = 43; S = score selected as standard for the grade = 49.

In words the meaning of the table is that of 1,401 eighth-grade children measured as to their knowledge of the multiplication tables, 2 could write answers at a rate of but 5 answers a minute,

TABLE I

DISTRIBUTION OF INDIVIDUAL SCORES IN SPEED TEST, MULTIPLICATION

Three Representative Classes; also Total of Fifty Classes; Eighth Grade

	No. in Class	Scores											
		0–9	10–19	25	35	45	55	65	75	85	95	105	115
Highest	36	...	.....	..	4	5	6	13	4	4	..	...	...
Median	43	...	.....	4	18	16	3	2	..	..	..	...	...
Lowest......	13	...	1	6	5	1	..	..	..	..	..	...	...
Total	1,401	2	17	180	460	410	191	77	26	22	5	10	1

17 at the rate of 15 answers a minute, 180 at 25 per minute, and so on, the average for the entire group being 43.

It is particularly to be noted that the wide range of the distribution is caused not so much by great differences between schools as by great variation in the abilities of individual children. The amount of variation in the grade averages of different schools does not equal the variation in the scores of the members of any one class. Table II and Fig. 2 show this plainly. The distribution of an average class of 47 members (selected from the 9 classes of the 51 having a membership of from 42 to 51) is given for comparison with the distribution of the 51 grade averages. The results are from the scores in the addition speed test, for many schools the ability in which the *least* variation was shown.

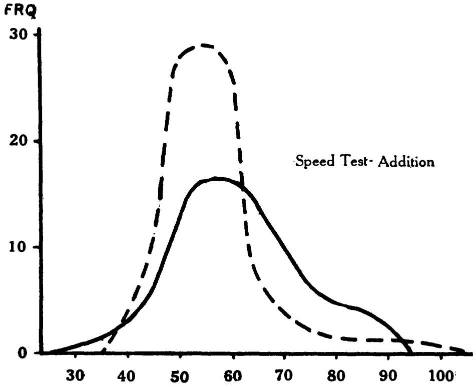


FIG. 2.—Comparison of the distribution of fifty eighth-grade averages with that of forty-seven individual scores in a single eighth-grade class. Solid line, individuals; broken line, classes. The curves show that the individuals in a single class differ more than the grade averages of the fifty schools.

TABLE II

SPEED TEST, ADDITION: EIGHTH GRADE

Comparison of Distribution of 47 Individual Scores in a Single Class of Average Variability with the Distribution of the Grade Averages of 51 Classes

	No.	Scores						
		30–39	45	55	65	75	85	95
Individuals	47	1	6	16	13	7	4	. .
Classes	51	. .	11	29	7	2	1	1

The fact shown is a constant feature of the returns for all the tests and grades, and can only mean that the differences in the abilities of individual children are greater factors in determining relative rank in school work than all the differences in abilities of teachers, courses of study, or methods of work combined. The result is that the grades in all the schools overlap to an extent well-nigh incredible, and thus the data from this more extended investigation fully confirms the conclusions reached in a discussion of this point in a previous article, although in that case the data was derived from a single school.

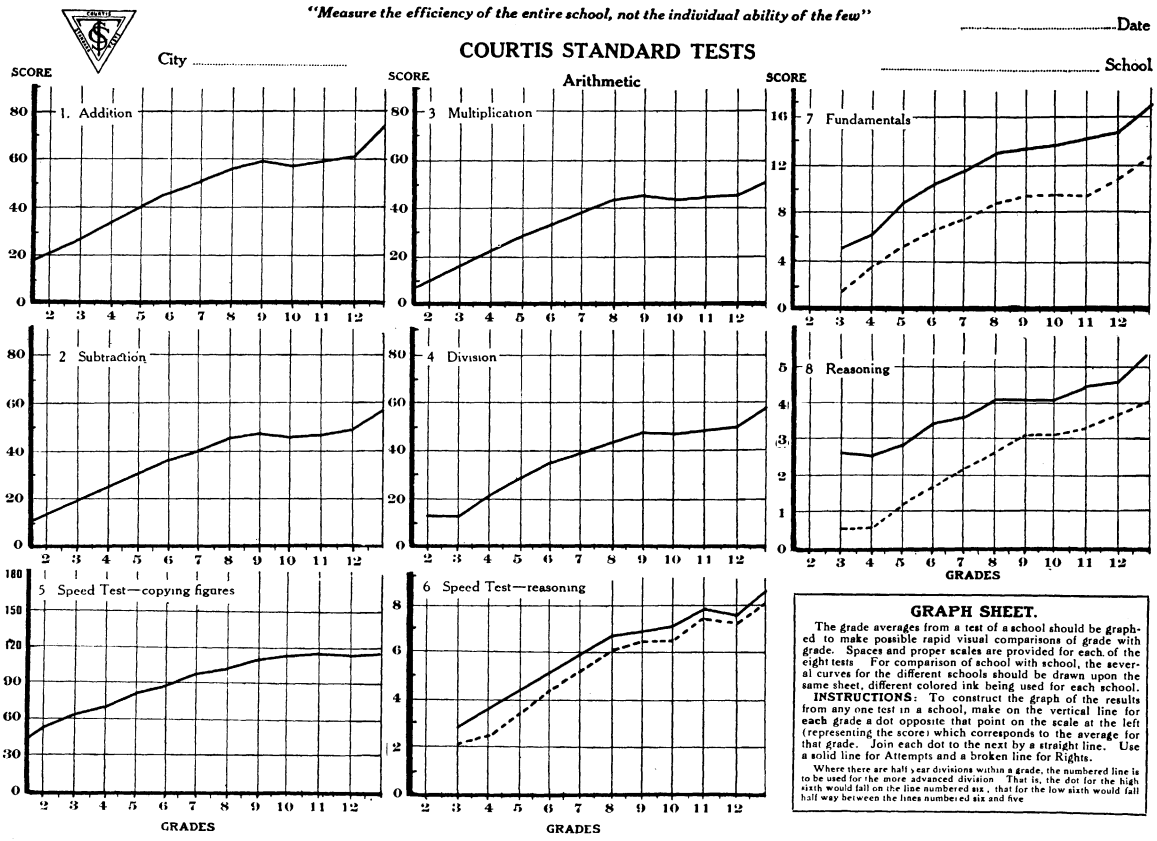
TABLE III

GRADE AVERAGES FROM TOTAL DISTRIBUTIONS

GRADE NO.	AVERAGE OF SCORES FOR EACH TEST	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6		No. 7		No. 8	
							Ats.	Rts.	Ats.	Rts.	Ats.	Rts.
1..............	55	6	6	..	..	29	...	...	....	....	...	...
2..............	75	21	12	10	12	51	...	...	....	....	...	...
3..............	525	26	19	16	11	63	2.8	2.1	5.4	1.7	2.7	0.6
4..............	1,222	33	25	23	21	70	3.7	2.5	6.6	3.6	2.6	0.8
5..............	1,177	40	32	30	28	80	4.4	3.4	9.0	5.3	2.8	1.2
6..............	1,282	46	37	34	35	88	5.1	4.4	10.3	6.9	3.4	1.7
7..............	1,432	51	40	38	38	98	5.9	5.2	11.5	7.6	3.7	2.2
8..............	1,370	57	45	43	44	102	6.8	6.1	13.1	8.9	4.1	2.7
9..............	412	59	47	45	47	108	6.9	6.4	13.7	9.5	4.1	3.1
10..............	216	57	45	43	46	112	7.2	6.7	14.0	9.5	4.1	3.1
11..............	151	59	47	44	48	114	7.9	7.4	14.4	9.4	4.5	3.3
12..............	169	61	48	44	49	112	7.7	7.2	14.9	10.8	4.6	3.6
13..............	462	71	56	50	56	116	8.6	8.2	16.8	12.6	5.3	4.0
14..............	131	74	51	58	59	124	9.7	9.1	17.2	11.8	5.4	4.1
	8,679*											

* Certain of the tests were omitted by several schools, thus materially reducing the average. For most of the tests, however, the total number of scores is over 9,000.

From the discussion above it is evident that the average for any grade represents simply a central tendency from which individuals vary widely. Yet in many ways the average scores of each group, given in Table III, and shown graphically in Fig. 3, reveal the character of the development of each of the abilities through the school and show what should be true of the development of the individual. These data, then, in connection with the distributions from which the averages are derived furnish the desired basis for the selection of standard scores.



GRAPH SHEET.

The grade averages from a test of a school should be graphed to make possible rapid visual comparisons of grade with grade. Spaces and proper scales are provided for each of the eight tests For comparison of school with school, the several curves for the different schools should be drawn upon the same sheet, different colored ink being used for each school.

INSTRUCTIONS: To construct the graph of the results from any one test in a school, make on the vertical line for each grade a dot opposite that point on the scale at the left (representing the score) which corresponds to the average for that grade. Join each dot to the next by a straight line. Use a solid line for Attempts and a broken line for Rights.

Where there are half year divisions within a grade, the numbered line is to be used for the more advanced division That is, the dot for the high sixth would fall on the line numbered six, that for the low sixth would fall half way between the lines numbered six and five

FIG. 3.—Grade averages. Solid lines, attempts; dotted lines, rights. The curves show for each ability the general character of its development through the school. Each curve is based upon approximately 9,000 individual scores.

This table means when put in words that in the third grade, for example, there were 525 individual scores in each test—or about 525. If in some test there were more, as for example, addition, there were enough less in division to make the average number of tests 525. In the addition tests it was found that the standard or average ability was that exhibited by a child who could record 26 in a minute. The standard in subtraction was 19 per minute, and so on.

The writer regrets exceedingly that there is not space to print the distributions in full; for the selection of standard scores must for the present be a matter of individual judgment. Many will be content to use the average scores as standards, and if, indeed, no children in any grade had a lower score than the present average for that grade, the increase in the efficiency of our schools in these abilities would be very great. But the writer feels that a higher score than the average is desirable for a standard and the scores given in Table IV were derived as follows: For each test an eighth-grade score was selected such that it was equalled or exceeded by thirty per cent of the eighth-grade children measured. This score was plotted and a smooth curve drawn, having the same general form as the average curve and coinciding with it in the lower grades. The scores for each of the other grades were then determined from the graph.

TABLE IV

STANDARD SCORES

TEST NO.	NO. 1	NO. 2	NOS. 3 AND 4	NO. 5	NO. 6		NO. 7		NO. 8	
					Ats.	Rts.	Ats.	Rts.	Ats.	Rts.
Grade 3	26	19	16	58	2.7	2.1	5.0	2.7	2.0	1.1
Grade 4	34	25	23	72	3.7	3.0	7.0	3.3	2.6	1.7
Grade 5	42	31	30	86	4.8	4.0	9.0	4.9	3.1	2.2
Grade 6	50	38	37	99	5.8	5.0	11.0	6.6	3.7	2.8
Grade 7	58	44	44	110	6.8	6.0	13.0	8.3	4.2	3.4
Grade 8	63	49	49	117	7.8	7.0	14.4	10.0	4.8	4.0
Grade 9	65	50	50	120	8.6	7.8	15.0	11.0	5.0	4.3

While the results given above were obtained by the use of special tests, and while for exact comparison the reproduction of the identical tests and the identical conditions are necessary, the writer hopes that teachers generally will appreciate the fact that the scores above furnish definite objective standards toward which to work. Accurate measurements are needed twice a year; once at the beginning for purpose of diagnosis, and again at the close of the year to measure the effect produced by the teaching effort. But for general class work, approximate values are all that are required. The scores in Table IV should be translated into words: thus, at the end of a year's work, an eighth-grade child should be able to copy figures in pencil on paper at the rate of 117 figures a minute; to write answers to the multiplication combinations at the rate of 49 answers per minute; to read simple one-step problems of approximately 30 words in length and decide upon the operation to be used in their solution at the rate of 8 examples a minute with an accuracy of 90 per cent; to work abstract examples of approximately 10 figures (twice as many for addition) at the rate of 14.4 examples in 12 minutes with an accuracy of 70 per cent; to solve two-step problems of approximately 10 figures at the rate of 5 in 6 minutes with an accuracy of 75 per cent. At the present time 70 per cent of the eighth-grade children cannot meet these standards. But it must be borne in mind that 3 per cent of the fifth-grade children can, and that experience has shown that individual care and a very little, well-managed drill produces marked changes in the abilities of most children.

A word of caution is necessary for those who use these standards. The writer is well aware of the disfavor into which drill has fallen, and he is heartily in sympathy with the ideas and principles which lead to its disuse. On the other hand, no fair-minded person can consider the wide range of variation in all the classes and not realize that something is wrong. It may well be that the pendulum has swung too far. Inordinate drill is surely harmful. Drill out of place is also harmful. But drill in its proper place—after, and not *before*, understanding control—and in proper amount—enough to insure the minimum ability necessary to effective work but not enough to involve waste of teaching effort—such drill is one of the vital factors in efficient teaching. The scores above, while necessarily tentative only, have been selected with due regard for such critical points in the development of these abilities as have been determined. It is probable that development much beyond the points indicated by the standard scores—and such development is possible—would involve waste of effort and, if not positively harmful, would be useless from the standpoint of efficient, general development. Accurate determination of the proper amount and form of drill required to produce the standard abilities, and similar determinations of standard, yearly growths, are the next problems that will be attacked. For the writer considers that the present investigation has proved that co-operative work on such educational problems is both feasible and of value.

The writer cannot bring this discussion to a close without a plea for both the teacher and the child. When a child's score exceeds the standard for his grade, he will ordinarily need little further attention. The ordinary procedure in grade work furnishes him adequate opportunities and incentives. But when the score is below the standard the case is entirely different. The fault may be in the child. A defective mentality may render the development of a given ability absolutely out of the question. But in such cases it would seem the height of folly and a rank injustice to give a teacher a fourth-grade intellect upon which to work and expect him to secure eighth-grade achievement. Or the child may be normal, but some peculiarity of its mental equipment causes a marked bias in one direction or another. The child that learns addition readily, but falls far below the standard in multiplication, needs special care and special work until the standard ability in multiplication is attained. A balanced, many sided development is the right of all. It is in such cases that standard scores are of the greatest service. On the other hand, the fault may be in the teaching, the course of study, or the methods of work. But here again the number of children who reach, or fail to reach, the standard is the measure of efficiency. From every point of view, therefore, standardization of all school subjects, and the use of standard scores offer a possible escape from present, unsatisfactory conditions.